图书情报工作
第 64 卷 第 7 期 2020 年 4 月

我国研究生在科研活动中的数据素养现状研究
——以双生命周期理论为视角

■ 余维杰 周娅莉 吴锦池
中山大学资讯管理学院 广州 510006

摘 要：[目的/意义]基于双生命周期理论视角，调查我国研究生在科研活动中的数据素养现状，并针对调查分析结果给出提升我国研究生在科研活动中数据素养的相关建议。[方法/过程]通过梳理国内外数据素养的概念和内涵，初步构建出数据素养的评价体系。在此基础上，根据数据生命周期和科研生命周期之间的嵌入关系，进一步构建出基于双生命周期下的研究生数据素养评价体系，以该体系中的目标技能作为参考编制出调查问卷，向全国不同高校、专业和年级的研究生发放问卷，并对调查结果进行分析。[结果/结论]调查结果发现，我国研究生具有良好的数据道德和法律意识，但是数据能力普遍较弱，深层次的数据意识还有待提升，不同学科、年级的数据素养具有一定差异。同时，针对该调查结果，提出对应的解决措施。

关键词：数据素养 科研生命周期 数据生命周期 双生命周期
分类号：G252
DOI：10.13266/j.issn.0252-3116.2020.07.010

M. Sheilds 在其 2004 年发表的论文 *Information literacy*、*statistical literacy*、*data literacy*[1] 中首次提出数据素养概念以来，为顺应大数据时代的发展，国内外各高校和相关机构都在大力宣传和推进数据素养教育。研究生作为科学研究的专业人才，数据素养已经成为其必备的核心素养之一。

目前，在众多学者的努力下，围绕着如何提升学生的数据素养能力的新型教学模式正逐步形成。国内外学者大多针对在科研生命周期和数据生命周期中的单个周期来探究数据素养的教育实践和理论研究[2-3]。随着研究的深入，关于数据素养与科研生命周期和数据生命周期三者之间关系的研究也逐渐走入学者的视线，学者们意识到从研究项目启动之初的数据产生、项目实施中的数据处理和分析，到项目结题时基于数据得出研究结论、发表科研成果，再到后续研究中对数据的二次开发和利用，科研数据已成为科学研究工作的基本单元渗透到科研工作流的各个环节[3]。从双生命周期的重要性来看，基于双生命周期的研究生数据素养研究将传统对数据素养的总况研究，细化到科研和数据生命周期的各个阶段中，能够更加精确、全面地对研究生数据素养的现状进行评价，对促进研究生数据素养相关工作具有指导性意义。因此，研究生数据素养培养工作的周期性研究是非常必要的。

笔者通过建立双生命周期下的研究生数据素养评价体系，探究基于双生命周期的研究生数据素养现状。并以此为依据，以期为研究生数据素养教育的实践提供参考和建议。

1 研究现状

从数据素养的内涵来看，国内外学者对数据素养的理解各抒己见。2007 年 E. Stephenson 等指出，数据素养是有效而恰当地发现、评估与使用信息和数据的一种意识和能力[4]；国外学者 K. Hogenboom 等对数据素养进行充分的研究并定义其能力内容包括阅读、解释、分析、批判性思考数据以及将数据作为证据的一种综合能力[5]；P. J. Calzada 认为通过程序化设计和自我评价，并结合有效数据来源、类型、获取途径等方面的特征，确定合适的方法来操作、分析、定量展现数据，以数据为主体的方式去处理问题，在解决问题的过程则能够体现数据素养能力[6]。近年来，国内的专家学

作者简介：余维杰（ORCID：0000-0002-0605-2296），副教授，博士，硕士生导师，E-mail：yuweijie6@mail.sysu.edu.cn；周娅莉（ORCID：0000-0001-9982-7953），硕士研究生；吴锦池（ORCID：0000-0002-8859-7950），硕士研究生。
收稿日期：2019-08-04 修回日期：2019-10-11 本文起止页码：84-93 本文责任编辑：徐健

者们也相继对数据素养进行研究。2013年，马云鹏站在小学数学教育的角度，提出在对当代小学生的数据素养培养过程中必须重视学生对数据的收集和统计能力。此外，还需要培养学生对数据的来源、数据形式进行质疑和思考的能力[7]。同年，张静波针对数据素养开展研究时，针对科研过程中涉及的数据获取、整理、分析等流程中应该具备的技能进行分析。此外，她还认为科研过程中研究人员在处理数据时应该遵守基本的道德规范，在数据生命周期中实现对数据的有效使用[8]。

2015年，美国大学与研究图书馆协会（Association of College & Research Libraries，ACRL）机构通过了美国高等教育的信息素养框架文件[9]，该框架将数据素养视为学术过程的一部分，将两者进行了融合，但又特别强调了将数据素养教育融入学术活动的生命周期后，数据素养教育仍然要注重周期性和嵌入性，因为数据素养教育也有其自身的生命周期。2016年，美国弗吉尼亚大学图书馆开展了"嵌入科研工作流与数据生命周期的数据素养能力研究"的实践，该项实践主要是以美国弗吉尼亚大学图书馆在官方网站的嵌入科研工作流的数据生命周期模型为导向进行的。具体来讲就是图书馆官方网站采用科研生命周期导航的方式，针对不同科研阶段的用户提供专门的数据信息服务[10]。康奈尔大学图书馆、伦敦国王学院图书馆和悉尼大学图书馆也纷纷效仿美国弗吉尼亚大学图书馆为研究者提供整个科研生命周期的服务，如高水平的研究咨询、数据分析、学术交流、知识产权及出版等服务[11]。近年，国内学者也开始将生命周期理论引入数据素养的研究当中。郝媛玲等提出，数据密集型科研范式的兴起，使得高校的科研工作者进行科研数据收集、管理、出版、引用等工作的难度越来越大，因此科研工作对科研工作者处理科研数据的能力要求会越来越严格，特别是科研工作者对待数据的态度、分析数据的技能等方面的数据素养已经渗透到科学研究的生命周期之中，影响科研项目的各个阶段，往往会直接影响到科研工作的好坏[12]。张军以科研项目作为主体，构建了基于科研数据生命周期理论的科研人员数据素养能力培养框架[13]。胡卉等梳理了嵌入科研工作流和数据生命周期的数据素养核心内容体系，绘制了研究人员应具备的数据素养能力框架[3]。

目前，学者们大多以其中的单个生命周期结合数据素养进行研究。而科研生命周期和数据生命周期都是以数据的操作和价值实现为目标的周期，两者既相互统一，又相互交叉。从这个角度看，基于双生命周期结合数据素养进行研究，能够丰富数据素养的研究维度，更加精确、全面和科学地对数据素养进行研究。因此，基于双生命周期建立数据素养评价体系，为探究研究生数据素养的差异现状和原因提供了新的视角。基于该探究结果可为研究生数据素养教育提出建议，对提升研究生数据素养水平具有一定的现实意义。

2 研究设计

本研究的研究对象为我国研究生科研活动中的数据素养，包括来自不同学校、专业和年级的高校研究生。采用调查问卷法，结合双生命周期模型，编制了基于双生命周期下的研究生数据素养调查问卷，采用线上和线下同时发放问卷的形式，根据实际的硕士、博士生比例，科学地配比发放。

本文的研究思路是在对国内外基于双生命周期和数据素养的研究成果进行介绍的基础上，结合国内研究生群体的特征，编制针对双生命周期下的研究生数据素养现状调查问卷。紧接着，分别从数据素养整体和双生命周期视角，分析总结研究生群体的数据素养情况，并提出提高研究生数据素养的建议。

3 模型与评价体系构建

3.1 双生命周期模型

2006年，学者J. Humphrey提出知识转移模型，该模型由6个阶段组成：概念化、初始化（开始阶段）、分析、产生初始结果、正式化和推广。该模型为科研生命周期的建立奠定了基础。此后，国内外不少学者都对科研生命周期进行了研究，如英国联合信息委员会认为科研生命周期应该包含5个周期：形成概念期、寻求合作期、写作期、研究期、出版期[14]。李文文则将科研生命周期分为4个阶段，分别是科研启动阶段、科研计划阶段、科研实施阶段、成果发表阶段[15]。北卡罗来纳州州立大学图书馆将科研生命周期分为了5个部分，分别是产生想法期、资金寻求期、建议期、持续进行期和传播期，根据不同时期的服务特点为科研人员提供相应的保障[11]。笔者参照研究生在科研生活中经历的科研阶段，更加倾向于将科研生命周期分为4个阶段，分别是科研启动期、科研准备期、科研实施期和成果发表期。

对于数据生命周期的模型，目前已有官方权威机构和学者提出了一系列的模型结构，英国数据档案机构建立了针对科研人员、数据密集型群体的环形生命

周期循环图,并对从“研究计划阶段 – 收集数据 – 处理和分析数据 – 发表和分享数据 – 保存数据 – 数据的再次利用”中涉及的每个数据生命阶段中需要掌握的数据能力进行了详细介绍[16]。英国联合信息委员会针对大学研究人员和研究管理支持人员,建立了一个闭锁环形的数据生命周期循环图,将数据生命周期定义为“从数据创建和存储 – 数据管理 – 数据库及档案 – 数据目录和注册计划和设计 – 收集和获取 – 合作和分析 – 管理和储存 – 分享和发布 – 发现和再次利用的过程[17]”。国际标准协会数据文档倡议联盟发布了 DDI (Data Documentation Initiative)数据生命周期模型,该模型与其他机构的生命周期模型存在差异,其采用的是半循环式的结构,认为数据生命周期应该包括“概念 – 集合 – 处理 – 分布 – 发现 – 分析 – 再利用 – 处理 – 存档 – 分布”几个步骤[18]。相较其他模型,英国数据档案(UK Data Archive, UKDA)模型的内容构成简洁、全面,具有普适性;在模型的呈现上,笔者认为研究生日常的数据使用流程,更贴近 DDI 提出的半循环式的生命周期,即并不是每一个部分都需要参加生命周期的循环。以此两种生命周期为参考,笔者构建了数据生命周期模型(见图 1)。

在研究生的科研生活中,科研生命周期与数据生命周期有着密不可分的关系。科学研究工作本身就是对数据进行探索,科研活动中嵌入大量的数据工作,科研数据的累积又推动了科研活动的发展。两者联系紧密,又相互促进。基于此,有关科研生命周期与数据生命周期的双模型逐渐成为了研究趋势,目前已分别有学者从二者的需求关系[16]、内容共性[19]、交融模式[20]等方面进行深入的研究。

由于针对双生命周期这一概念的研究时间较短,因此还处于初见雏形的阶段。通过对以上所述国内外当前在数据生命周期和科研生命周期模型方面的研究状况进行全面的分析和整合的基础上,综合国内外相应的研究实践内容,最终构建出双生命周期模型框架(见图 1)。

图 1 中的双生命周期是在数据生命周期和科研生命周期的基础上建立的,不同方框表示不同的生命周期阶段,同时也反映出数据生命周期嵌入科研生命周期的情况,以此来解释两个生命周期之间的关系。

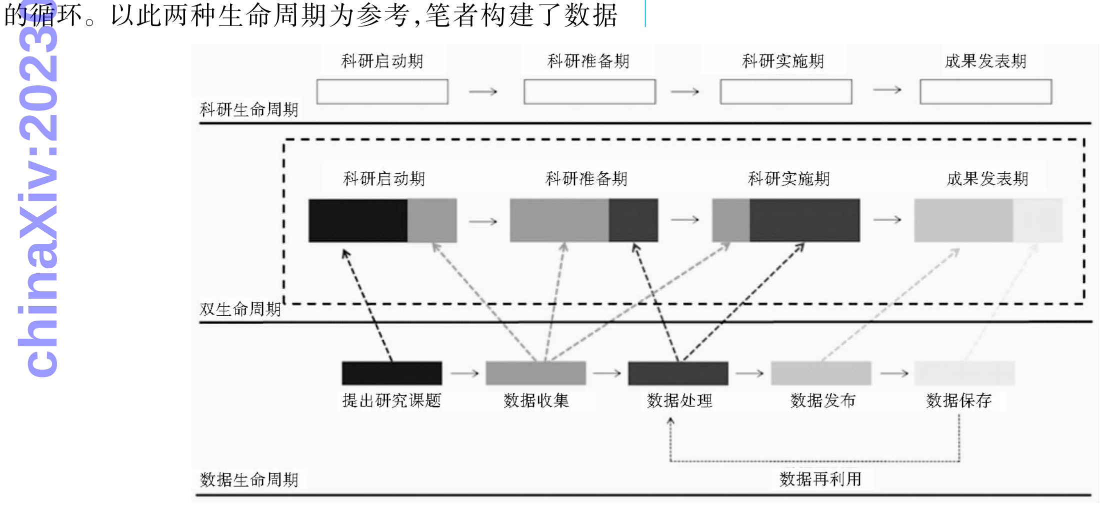


图 1 双生命周期模型框架

在科研生命周期中的科研启动期是科研的开端,该阶段主要以意识为主导,涉及简单的数据收集能力。在科研启动期确立好研究方向之后,就进入科研准备期,科研准备期通过收集大量的数据,并加以分析,以此来为科研实施做好准备,科研准备期主要对应的数据周期包括数据收集和数据处理阶段。科研实施期作为一个承上启下的阶段,主要包括科研实验和文章撰写两部分,科研实验通过整理实验中得出的数据和获取外部数据,再经由一系列的数据处理手段得出科研结论,最后将实验结论通过可视化或者文字形式撰写出来。科研实施期主要对应数据生命周期阶段中的数据收集阶段和数据处理阶段。最后的成果发表期,需要研究者将数据发表到与研究方向相对应学科的期刊、数据库中,因此对应数据生命周期末端数据的发布与保存。数据保存中除了保存最终发表的数据,在实验中获得的未发表的数据也应进行整理、保存。

3.2 数据素养评价体系

随着研究者们对数据素养研究的深入,数据素养评价体系的构建为国内外研究者争相探索。E. S. Gummer 等人基于领域分析方式构建了数据素养评价

框架，该框架中的数据素养评价指标包括确认问题、提出问题、使用数据、将数据转换成信息、将信息转换成决策、结果评价6个部分[21]。S. Z. Athanases等基于教师行业构建了教师数据素养评价体系，通过充分探究教师行业从业人员的数据素养需要，该研究中的数据素养评价指标包括数据操作能力、阐述问题能力、收集和整理数据能力、数据分析工具使用能力和呈现数据能力[22]。以上学者所构建的数据素养评价体系均侧重于数据能力，而曹树金等则认为数据素养不能仅强调数据能力，还应该包括数据意识和数据伦理[5]。笔者在参考借鉴相关研究成果后，认为综合型的数据素养组成类型对数据素养的解读更加客观全面，因此更认同综合型的数据素养组成类型，即数据素养由数据意识、数据能力和数据伦理3个部分组成。

数据意识指的是个人对数据信息的敏感程度，同时也是数据指导下的综合行为[23]。鉴于此，可以将数据意识定义为在大数据环境下，数据使用者能够清晰认知自身数据使用需求的一种意识。数据能力是一种实践型的能力，贯穿数据的获取、整理、分析、保存整个数据生命全周期。数据能力作为数据素养中核心的部分，是研究生在整个科研生命周期中反复进行的。数据能力的高低除决定开展科研工作的可能性以外，直接影响最终处理结果的好坏。数据伦理指能够合理合法地获取并使用数据，在数据的发布和使用过程中能遵守相关社会法规与规范，采用正确的使用方法，并坚持伦理性原则。如今的大数据时代，数据泛滥，人们获取数据的途径变多、获取数据的成本也变低，但随之而来的数据侵害、数据滥用等有违伦理道德的事情屡见不鲜。加强人们的数据伦理，保证数据能够在合理合法的情况下使用和传播，在大数据时代这点显得尤为重要。

综上所述，笔者认为在数据素养的3个组成部分当中，数据意识决定科研活动中的视野与高度，数据能力决定研究内容的深度和广度，数据伦理决定研究过程中行为的规范。

目前，国内外专家学者对于数据素养评价体系和基本框架尚未达成共识。笔者在综合大量有关数据素养评价体系和基本框架的文章并参考了2015年美国权威发布的《高等教育信息素养框架》后[24]，进一步结合数据生命周期和科研生命周期的特性和研究生科研活动实际状况，建立了数据素养评价体系（见图2）。该体系中的一级指标由数据意识、数据能力和数据伦理构成，即前文中探讨出的数据素养的3个组成部分。其中，数据意识包含获取意识、使用意识和传播意识3个二级指标；数据能力则包含数据收集能力、数据处理和分析能力、数据展示能力、数据保存能力4个二级指标；数据伦理包含数据规范、数据安全和数据法律3个二级指标。在上述10个二级指标下，又进一步划分了19个目标技能。

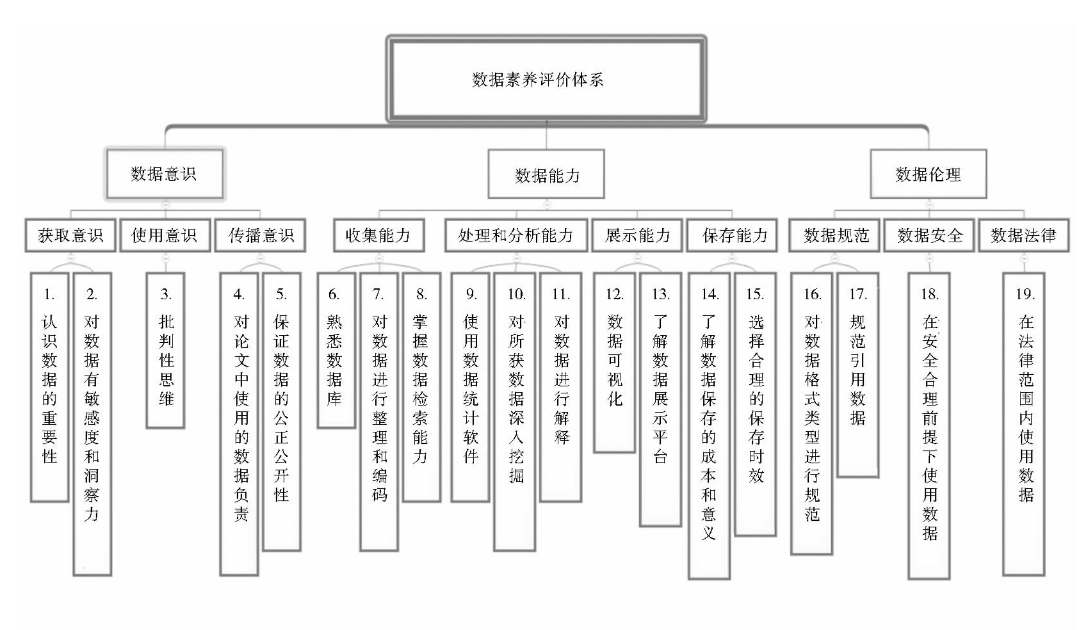


图2 数据素养评价体系

3.3　基于双生命周期理论的数据素养评价体系

目前数据素养教育的开展情况中，可以从横向和纵向对数据素养教育进行概括。从纵向角度来看，数据素养从数据意识、数据能力、数据伦理等角度进行逐步分层展开，即从数据意识的培养逐步上升到对数据的收集、应用、存储等能力层面，该层面站在科学数据的资源导航、通识教育、学科素养等科学数据的角度，以数据生命周期为主线进行数据素养的教育。从横向角度来看，数据素养教育站在科研生命周期的角度，从科研启动期到成果发表期来探讨数据素养教育的开展。

正如前文提到的双生命周期是以科研生命周期为主线，数据生命周期嵌入科研生命周期的模型。笔者以图 1 中的双生命周期模型和图 2 中的研究生数据素养评价体系为基础，结合各阶段中的数据素养需求特征，构建出基于双生命周期的数据素养评价体系，如图 3 所示：

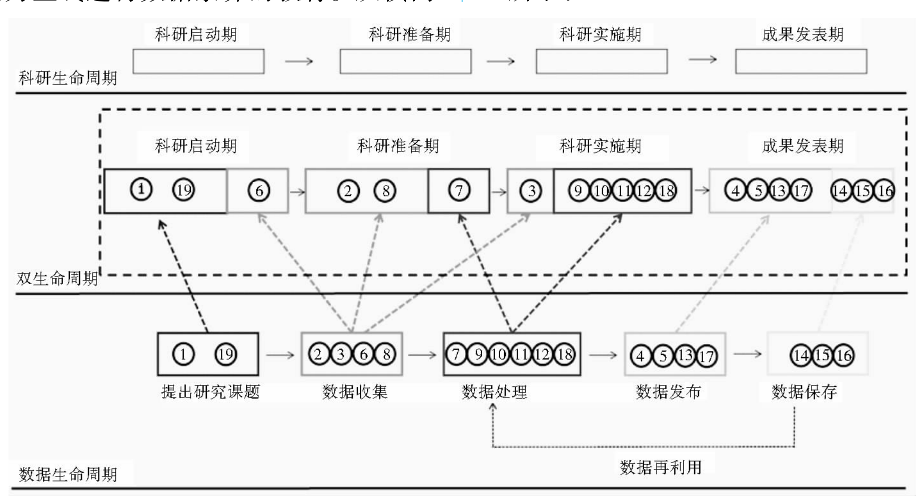


图 3　基于双生命周期的数据素养评价体系

该基于双生命周期的数据素养评价体系不仅体现了两个生命周期之间的嵌入关系，同时用不同数字对应数据素养评价体系中的目标技能（如数字 1 对应的是图 2 中的第一项目标技能“认识数据的重要性”），更加直观地展示了在不同生命周期中，不同阶段的数据素养技能需求。笔者将以此评价体系为核心设置问卷，并以问卷调查的形式探讨当前研究生的数据素养现状。

4　研究生数据素养现状分析

4.1　研究生数据素养调查设计

在调查对象选取上，本次研究的主体为双生命周期下研究生数据素养的现状探究，因此问卷调查的研究对象为研究生。采取随机抽样的方式，对研究群体进行抽样。在被调查者地域的选择上，为保障本次研究对象数据素养的高低不受地域的影响，笔者调查了来自我国北部：北京市、黑龙江省；中部：重庆市、四川省；南部：广东省在校研究生。在被调查者的学校选择上，为避免研究生数据素养能力水平会因为学校的不同而存在偏差，本次调查对“985”院校、“211”院校和普通院校的样本都有选择，保证实验的科学性。本次调查涉及的研究生群体有博士生和硕士生，其中对硕士生也进行了不同年级、专业型硕士和学术型硕士的分类。研究生学科包括人文社科类、自然科学类、工程与技术科学类、医学类、农科类和其他学科。基于李克特五级量表，本次问卷的计分模式按照每一道题从左到右，强度依次增强，每道题的计分也从 1 – 5 依次增加。为使呈现出的数据更加直观，根据在五级量表中得到的分数，笔者用百分比的形式来评估研究生的数据素养水平。按照总分的 85% 以上（包含 85%）、总分的 70% – 85%（包含 70%）、总分的 60% – 70%（包含 60%），及 60% 以下的分别设置为优秀、良好、合格及较差。

4.2　调查问卷信度和效度分析

本次样本共收集到数据 754 份来自不同地区的研究生问卷。删除回答不完整的和很明显胡乱填写的 4 份，本次共获得有效问卷 750 份，问卷回收率 99.47%。本次使用 SPSS 软件，对所获取的数据进行可信度和效度分析。剔除前面的关于被调查者信息类的题目，对关于数据素养现状的题目进行信度分析，得出的克朗巴哈（Alpha）系数为 0.862。该系数介于 0.8 – 0.9 之间，说明问卷信度非常好，可以进行下一步的深入分析。对获取到的数据进行效度分析，得到 Bartlett 球度系数为 0，当 Bartlett 系数小于 0.05 时，问卷具有结构

效度。因此本次的问卷具有结构效度。

4.3 研究生数据素养的整体现状分析

如前文所构建的数据素养评价体系，对各指标进行问卷题目设置。在数据意识方面，一共设置了5道题目，涵盖数据的获取意识、使用意识和传播意识3个二级指标；数据能力作为数据素养的最重要组成部分，一共设置了10道关于数据能力题目，分别涉及数据的收集能力、处理和分析能力、展示能力和保存能力4个二级指标；在数据伦理方面，设置了4道关于数据伦理的题目，涵盖数据法律、数据安全和数据规范3个二级指标。将有效问卷进行整理，计算每道题目平均得分，结果如表1所示：

表1　研究生数据素养整体现状

一级指标	二级指标	题目	得分	百分比(%)	等级
数据意识	获取意识	数据对科研的重要性	3.98	79.00	良好
		对数据的敏感度和洞察力	3.31	66.20	合格
	使用意识	对所获数据具有批判性的思维和习惯	3.25	65.00	合格
	传播意识	对发表数据负责	3.72	74.40	良好
		保证数据的公正性和开放性	3.76	75.23	良好
数据能力	数据收集能力	知悉获取相关数据库途径	3.13	62.69	合格
		对数据进行科学整理编码	3.05	60.99	合格
		检索数据能力	3.30	65.95	合格
	数据处理和分析能力	合理找到统计软件分析	3.33	66.61	合格
		清晰解释分析出的数据	3.16	63.28	合格
		对数据进行深入挖掘	3.12	62.40	合格
	数据展示能力	清楚了解数据的呈现形式	3.37	67.33	合格
		了解数据的对应展示平台	3.21	64.13	合格
	数据保存能力	了解保存数据的成本意义	3.22	64.43	合格
		了解数据的价值和合理的保存时效	3.39	67.89	合格
数据伦理	数据规范	尊重他人的数据，规范地对引用的文献能够标明出处	4.01	80.20	良好
		是否对数据格式、类型规范	3.69	73.84	良好
	数据安全	是否保障数据能够在安全、合理的条件下使用	4.07	81.40	良好
	数据法律	是否在法律和道德的允许范围内使用数据	4.39	87.87	优秀

从表1中可以看出，研究生整体的数据意识中等偏上。单独来看，大部分研究生都很清楚数据对科研的重要性，该题目在数据意识中所得到的分数是最高的，同样表现良好的是学生们对数据的负责和保证数据的公正性和开放性，均表现良好态势。但是在对数据的批判性思维和对数据的敏感度和洞察力等进一步较为深入的意识探究上，研究生的表现就稍次一些，仅是合格水平。以上也说明了，研究生的数据意识仅停留在一些较为表面的层次上，而对数据进行深度挖掘、深度思考的意识较弱。

数据能力是数据素养教育中最为核心的部分。如表1所示，在数据素养的所有数据能力中，数据保存能力得分最高，其次是数据展示能力，紧接着是数据处理和分析能力，数据收集能力得分最低。可以看出，研究生在数据能力上总体仍然比较欠缺。特别是得分最低的数据收集能力上，对于收集到的数据，能够对其进行科学整理和编码能力欠佳，仅是及格线水平。良好的科学整理和编码能力能够使后面的数据调查和科研输出事半功倍，因此科学的数据整理与编码能力是研究生数据素养教育中是不容忽视的一种能力。在被调查的研究生们对于数据的深入挖掘能力也较弱，数据挖掘是一个非常具有应用价值的新领域，但由于其技术难度较大，研究生很难在日常的科研活动中广泛展开应用。同样得分较低的还有知悉获取相关数据数据库的能力。

数据伦理是每个人在数据活动中需要遵守的基本准则。从表1中可以看出，大多数研究生对于数据法律会遵守数据在使用上的法律和道德、能够安全地使用数据、在数据发表时能够规范地标注引用的文献，但在对数据格式和类型规范上，相对其他数据伦理则较弱。这也从侧面说明，研究生们对数据法律、数据安全等数据伦理中的红线问题很注重，但对数据的规范较为轻视。

4.4 基于双生命周期的研究生数据素养现状分析

本次通过科研生命周期为主线进行问卷设计，在前文构建的基于双生命周期的研究生数据素养能力框

架的基础上，分析出基于科研生命周期下研究生数据素养现状的同时，还以此推导出研究生在数据生命周期下的数据素养现状。探究双生命周期下研究生的数据素养现状，对于更好培养研究生在数据意识、数据管理等能力上具有重要的意义。

图 4 是基于双生命周期模型所探讨出的数据素养现状图。其中，左边表格是科研生命周期与贯穿其中的数据素养目标技能，右边表格是数据生命周期与其对应的数据素养目标技能。两个生命周期数据素养现状表中间用虚线连接，更加直观地表示了目标技能在双生命周期中的对应关系，同时也体现两个生命周期之间的嵌入关系。

总得分	得分	等级	二级指标	科研生命周期
3.83	3.98	良好	获取意识	科研启动期
	4.39	优秀	数据法律	
	3.13	合格	数据收集能力	
3.22	3.31	合格	获取意识	科研准备期
	3.3	合格	数据收集能力	
	3.05	合格	数据收集能力	
3.38	3.25	合格	使用意识	科研实施期
	3.33	合格	数据处理和分析能力	
	3.16	合格	数据处理和分析能力	
	3.12	合格	数据处理和分析能力	
	3.37	合格	数据展示能力	
	4.07	良好	数据安全	
3.57	3.72	良好	传播意识	科研发表期
	3.76	良好	传播意识	
	3.21	合格	数据展示能力	
	4.01	良好	数据规范	
	3.22	合格	数据保存能力	
	3.39	合格	数据保存能力	
	3.69	良好	数据规范	

数据生命周期	二级指标	等级	得分	总得分
提出研究课题	获取意识	良好	3.98	4.19
	数据法律	优秀	4.39	
数据收集	数据收集能力	合格	3.13	3.25
	获取意识	合格	3.31	
	数据收集能力	合格	3.3	
	使用意识	合格	3.25	
数据处理	数据收集能力	合格	3.05	3.35
	数据处理和分析能力	合格	3.33	
	数据处理和分析能力	合格	3.16	
	数据处理和分析能力	合格	3.12	
	数据展示能力	合格	3.37	
	数据安全	良好	4.07	
数据发布	传播意识	良好	3.72	3.68
	传播意识	良好	3.76	
	数据展示能力	合格	3.21	
	数据规范	良好	4.01	
数据保存	数据保存能力	合格	3.22	3.43
	数据保存能力	合格	3.39	
	数据规范	良好	3.69	

图 4　基于双生命周期的研究生数据素养评价

图 4 左边部分是根据研究生在科研生命周期中各阶段的数据素养得分，对应相应的评定等级制作的研究生在科研生命周期中的数据素养现状表。从研究生在科研生命周期下的数据素养现状可以看出，科研启动期的数据素养分数最高，其次是科研发表期，两个期间数据素养的等级评定都为良好；在科研实施期和科研准备期中，两者等级为合格。纵观研究生在整个科研生命周期中的数据素养得分，学生数据素养得分最高的是科研启动期，最差的是科研准备期。科研准备期作为研究者根据研究目的对资料进行收集、整合的阶段，与科研实施期的数据处理与数据结论的得出都存在着科研准备期存在较大的关系。

图 4 右边部分是根据研究生在数据生命周期中各阶段的数据素养得分，对应相应的评定等级制作的研究生在数据生命周期中的数据素养现状表。从在数据生命周期下的数据素养现状可以看出，研究生在提出研究课题阶段和数据发布阶段的综合目标技能等级都为良好，但在数据收集、数据处理阶段和数据保存阶段，其能力评定的等级仅为合格水平，数据收集阶段分数最低，数据处理阶段次之。从现状图中可以看出，数据处理阶段、数据收集阶段都是数据素养中数据能力密集型的两个阶段，而在评定中获得较高得分的提出课题阶段和数据发布阶段并非是数据能力密集型的阶段。在研究生的日常科研中，数据收集和数据处理阶段是科研成果成败的关键两个阶段。

4.5　基于双生命周期的研究生数据素养对比

4.5.1　数据素养水平与年级成正相关

为了探究不同年级研究生的数据素养现状，笔者根据年级不同，将研究生分为一年级、二年级硕士生、三年级硕士生和博士生 4 个大类。基于上文中所构建的双生命周期阶段划分方式，笔者从生命周期的每一阶段中对比出不同年级数据素养现状，探究数据素养水平高低与年级高低之间的关系，结果如表 2 所示（各项得分由每阶段对应题目得分平均求得）：

表2　基于双生命周期研究生数据素养年级对比

生命周期		研一	研二	研三	博士生
科研生命周期	科研启动期	3.69	3.80	3.80	4.20
	科研准备期	2.96	3.14	3.56	3.48
	科研实施期	3.17	3.39	3.60	3.54
	成果发表期	3.41	3.55	3.67	3.81
数据生命周期	提出研究课题	4.13	4.18	3.96	4.51
	数据收集	2.98	3.18	3.51	3.58
	数据处理	3.12	3.35	3.62	3.49
	数据发布	3.57	3.65	3.72	3.85
	数据保存	3.19	3.40	3.60	3.76

在科研生命周期中，各年级学生存在显著的差异。一年级硕士生到三年级硕士生，在科研生命周期中各阶段的数据素养水平都呈现出递增的现象。在科研启动期和科研发表期中，博士生的数据素养水平明显高于三年级研究生，而在科研准备期和科研实施期中，三年级硕士生与博士生之间存在的差异并不显著，甚至出现博士生数据素养水平略微低于三年级硕士生的现象。

在数据生命周期中，研究生的数据素养水平与其年级的高低呈现出的正相关性更明显。博士生与三年级研究生的数据素养整体水平相差不大；二年级研究生在各阶段的数据素养水平都高于一年级研究生；在提出课题阶段，三年级研究生的得分最低。除了三年级研究生提出课题阶段分数低于一年级研究生、二年级研究得分，博士生数据处理能力阶段得分低于三年级研究生外，研究生数据素养水平与年级高低总体呈现正相关性。

4.5.2　数据素养水平学科差异明显

不同学科间在知识属性、科研范式、学科文化和成果形式等方面存在明显差异，基于此，笔者将所有学科分为6个大类，分别为人文社科类、自然科学类、工程与技术科学类、医学类、农科类和其他类，来对数据素养在不同学科间研究生的现状水平进行比较，结果如表3所示（各项得分由每阶段对应题目得分平均求得）：

表3　基于双生命周期研究生数据素养学科对比

生命周期		人文社科类	自然科学类	工程与技术科学类	医学类	农科类	其他
科研生命周期	科研启动期	3.81	4.00	3.78	4.03	4.07	3.63
	科研准备期	3.15	3.36	3.17	3.45	3.41	3.07
	科研实施期	3.31	3.53	3.38	3.53	3.41	3.26
	成果发表期	3.49	3.77	3.58	3.65	3.63	3.43
数据生命周期	提出研究课题	4.20	4.33	4.18	4.22	4.44	3.93
	数据收集	3.19	3.40	3.17	3.53	3.44	3.09
	数据处理	3.26	3.50	3.35	3.51	3.37	3.25
	数据发布	3.63	3.81	3.70	3.72	3.69	3.51
	数据保存	3.32	3.71	3.43	3.56	3.56	3.32

在科研生命周期中，综合评分排名最靠前的是自然科学的学生，在科研实施期和科研发表期的评分都是最高的，其次是医学类和农科类的学生，而评分最低的为其他学科的学生。在科研生命周期下，倾向于理工科的研究生数据素养的评分大于文科的研究生，这也与研究生日常学习中的科研环境和科研需求有关。

在数据生命周期中，自然科学类的学生和农科类的学生数据素养的综合排名最靠前。对比以上学科在各数据生命周期中的差异，在提出课题阶段差异最大，各学科间的极差达到了0.52，跨度较大；以上学科中的研究生在数据处理阶段的评分极差为0.26，在该阶段的数据评分最为均衡。

综上可以看出，不同学科的研究生在数据素养水平上都呈现出较大的差异。具体来说，技术能力较强的理工科研究生，其数据素养的水平呈现出高于对数据能力需求较弱的文科研究生。差距的存在可能与不同专业研究生的数据素养水平有关，也可能与其自身专业在学习中的科研需求存在较大的关系。因此，在未来应该针对不同学科的学生，和其对数据素养需求和数据能力现状的不同，有针对性地开展数据素养教育工作。

5　提升研究生数据素养建议

5.1　加强开展基于双生命周期的研究生数据素养教育工作

国内的数据素养教育可以以国外的数据素养教育内容设计为参照，以科研生命周期为主线，在横向上，围绕科研周期和数据生命周期展开涉及数据收集、组织、管理、保存、利用与再利用的学习，以及相关政策、数据管理、数据共享、数据伦理等知识与技能的培训。在纵向上，通过数据素养通识教育以由浅入深的教育形式注重教育模式上的层次递进性[25]。如今图书馆作为在大学生数据素养教育中承担重要角色的施教单位，这也意味着高校图书馆将会为研究生提供包括数

据查询、数据共享等在内的更多培训。

在高校的日常教学中,根据数据素养的特征展开递进式的教育模式,为双生命周期下数据素养的提高起辅助作用。对研究生开展围绕科研流程的数据素养通识教育。从意识方面着手,提高数据需求、数据安全和数据伦理方面的敏感度,引导学生建立正确的数据价值观,善于利用数据提高专业技能[26];从数据使用角度着手,通过设置实践课程,对如数据资源的管理和建设、数据支持服务等的数据管理知识进行普及,让学生掌握数据使用的基本技能。

5.2 丰富以双生命周期为基础的数据素养教育形式

目前,我国高校的数据素养教育形式和内容设置上较为单一,主要以图书馆开设讲座为主,以培训和专项教育为辅,个别高校开设了相关的慕课课程,但都是在信息素养课程的子课程下进行学习的。以国外为借鉴对象,以美国伊利诺伊州香槟分校为例,该校在数据素养课程上设置了信息组织、伦理等多领域的数据管理教育课程[27]。

从调查结果可以看出,目前我国研究生的数据素养短板在于数据能力。未来在我国的数据素养教育上,高校可以通过组建科研小组的形式,学生通过置身于科研环境中,感受科研工作流中自身在数据管理中存在的问题和解决策略;高校提供多样的数据能力选修课程,学生可以根据自己所学的专业和自身数据能力的需求选择相应的课程,强化在自身学科科研过程中来提高数据能力;开展低学分的迷你课程,自主选择在生命周期中需要提高数据素养的阶段进行学习等。

5.3 推进多层次的研究生数据素养教育

对于不同年级、不同专业的研究生,他们的数据素养的现状和数据素养的需求不尽相同。因此,针对不同类型的研究生,应该因材施教,展开多层次的数据素养教育。

从研究生的年级高低与数据素养水平的高低呈现出的正相关关系可以看出,从事科研的时间长短和知识积累的多少是导致不同年级研究生之间出现数据素养水平差异的最根本原因。一年级研究生作为初融入科研的群体,应该重点加强数据意识和数据伦理的教育工作;而针对科研密集的高年级学生,应该重点展开数据收集和数据处理等数据能力型的教育。在数据素养的现状调查中,博士生群体与三年级研究生的数据素养水平差距并没有拉开,这与科研活动中博士生应该具备的数据素养不符合。在未来针对博士生群体的数据素养教育中,应该同步加强在数据意识、数据能力和数据伦理方面的教育,并且掌握的深度和广度都应该大于硕士生。

不同专业的学生由于研究领域的性质和自身知识结构存在的差异,在科研活动中的数据行为也各有特点。这一点在不同学科呈现出的研究生数据素养现状上也得以体现。理科生的数据素养水平总体高于文科生,数据密集型学科学生的数据素养水平总体高于数据使用较少学科的学生。针对不同专业学生的数据素养需求,可以借鉴加州大学洛杉矶分校的学科领域嵌入教学的合作方式,即专业授课老师与图书馆员共同合作进行数据素养教育[5]。

6 总结

总体来看,目前研究生对于数据的批判性思维和敏感度较低,深层次的数据意识还有待提升;研究生的数据能力普遍较弱,亟待加强;他们具有良好的数据道德和法律意识。从双生命周期的角度来看,研究生在科研启动期和科研发表期中数据素养良好,但在科研准备期和科研实施期中的数据素养水平偏低;在数据生命周期中,研究生在提出科研课题阶段数据素养评分最高,但在数据能力密集的数据收集期、数据处理期和数据保存期,研究生的数据素养仅为合格水平,仍有待提高。

本文旨在给数据素养教育的实施者和研究者提供参考,也给全国研究生数据素养的培养和提高提供一定借鉴。

参考文献:

[1] SCHIELD M. Information literacy, statistical literacy and data literacy[J]. IASSIST quarterly / International Association for Social Science Information Service and Technology, 2004, 28(2):7-14.

[2] 宋钧玉. 基于"互联网+"的高校图书馆数据素养服务平台探究[J]. 新世纪图书馆,2017(8):50-55.

[3] 胡卉,吴鸣. 嵌入科研工作流与数据生命周期的数据素养能力研究[J]. 图书与情报,2016(4):125-137.

[4] STEPHENSON E, CARAVELLO P S. Incorporating data literacy into undergraduate information literacy programs in the social sciences: a pilot project[J]. Reference services review, 2007, 35(4):525-540.

[5] 曹树金,刘慧云,王维佳. 从信息素养到数据素养[J]. 图书情报研究,2017(1):19-24.

[6] CALZADA P J, MARZAL M. Incorporating data literacy into information literacy programs: core competencies and contents[J]. Libri: international journal of libraries and information services, 2013,63(2):123-134.

[7] 马云鹏. 小学数学核心素养的内涵与价值[J]. 小学数学教育,2015(9):3－5.
[8] 张静波. 大数据时代的数据素养教育[J]. 科学,2013,65(4):29－32,4.
[9] 秦小燕. 美国高校信息素养标准的改进与启示——ACRL《高等教育信息素养框架》解读[J]. 图书情报工作,2015,59(19):139－144.
[10] Research data management [BE/OL]. [2019－06－28]. http://guides. library. vcu. edu/data.
[11] 薛晶晶,焦坤,张欣,等. 国外大学图书馆基于科研生命周期的科研支持服务研究[J]. 情报理论与实践,2016,39(5):110－114.
[12] 郝媛玲,沈婷婷. 数据素养及其培养机制的构建与策略思考[J]. 情报理论与实践,2016,39(1):58－63.
[13] 张军. 面向科研第四范式的科研人员数据素养培养研究[J]. 图书与情报,2016 (2):133－136.
[14] 索传军,王涛,付光宇. 国内外信息生命周期管理研究综述[J]. 图书馆杂志, 2008(7):14－20.
[15] 李文文,周萍. 基于科研生命周期的科学素养教育研究[J]. 现代情报,2017,37(1):156－159.
[16] UK DATA SERVICE. Research data lifecycle [EB/OL]. [2019－06－30]. https://www. ukdataservice. ac. uk/manage-data/lifecycle. aspx.
[17] LYON L. Dealing with data: Roles, rights, responsibilities and relationships [EB/OL]. [2019－06－30]. https://purehost. bath. ac. uk/ws/files/419529/dealing_with_data_report-final. pdf.
[18] Overview of the DDI version 3. 0 conceptual model [EB/OL]. [2019－07－02]. http://opendatafoundation. org/ddi/srg/Papers/DDIModel_v_4. pdf.
[19] 凌婉阳. 大数据与数据密集型科研范式下的科研人员数据素养研究[J]. 图书馆,2018(1):81－87.
[20] 丁宁,马浩琴. 国外高校科学数据生命周期管理模型比较研究及借鉴[J]. 图书情报工作,2013,57(6):18－22.
[21] GUMMER E S, MANDINACH E B. Building a conceptual framework for data literacy[J]. Teachers college, 2015, 117 (4):1－22.
[22] ATHANASE S Z, BENNETT L H, WAHLEITHNER J M. Fostering data literacy through preservice teacher inquiry in English language arts[J]. The teacher educator, 2013, 48 (1):8－28.
[23] 姚晓鹭. 大数据时代图书馆嵌入式学科馆员数据素养的构成与培养[J]. 农业图书情报学刊,2017,29(2):198－201.
[24] Association of College and Research Libraries. Framework for information literacy for higher education [EB/OL]. [2019－07－03]. http://www. ala. org/acrl/sites/ala. org. acrl/files/ content/issues/infolit/Framework_ILHE. pdf.
[25] 张群,刘玉敏. 面向研究生的高校图书馆科学数据素养教育研究[J]. 大学图书馆学报,2017,35 (3):93－98.
[26] 马晓慧. 大数据背景下大学生数据素养教育探讨[J]. 软件工程,2019,22 (2):48－50,41.
[27] 孟祥保,李爱国. 国外高校图书馆科学数据素养教育研究[J]. 大学图书馆学报,2014,32(3):11－16.

作者贡献说明:
余维杰:论文选题、研究方案设计,论文撰写与修改;
周娅莉:数据收集,论文初稿撰写;
吴锦池:数据处理与分析。

Investigation and Research on the Status of Chinese Postgraduates' Data Literacy in Scientific Research Activities ——From the Perspective of Dual Life Cycles

Yu Weijie Zhou Yali Wu Jinchi
School of Information Management, Sun Yat-sen University, Guangzhou 510006

Abstract: [**Purpose/significance**] Based on the double life cycle theory, this paper investigates the status quo of data literacy of postgraduates in scientific research activities in China, and puts forward some suggestions to improve the data literacy of postgraduates in scientific research activities according to the results of investigation and analysis. [**Method/process**] By combing the concepts and connotations of data literacy at home and abroad, this paper constructed a preliminary evaluation system of data literacy. On this basis, according to the embedding relationship between data life cycle and scientific research life cycle, a data literacy evaluation system for postgraduates based on double life cycle was further constructed. Questionnaires were compiled by referring to the target skills of the system. Questionnaires were distributed to postgraduates of different universities, majors and grades in China, and the results were analyzed. [**Result/conclusion**] The survey finds that graduate students in China have good data ethics and legal awareness, but data capabilities are generally weak, and deep-level data awareness needs to be improved. In addition, the data literacy of different disciplines and grades has certain differences.

Keywords: data literacy research life cycle data life cycle dual life cycle